Does Hypothesis-Instruction Improve Learning?

Regina Vollmeyer
Institut für Psychologie
Universität Potsdam
14415 Potsdam, Germany
vollmeye@rz.uni-potsdam.de

Bruce D. Burns
Department of Psychology
University of California, Los Angeles
Los Angeles, California 90095-1563
burns@psych.ucla.edu

Abstract

Dual space models of problem solving (e.g., Simon & Lea, 1974; Klahr & Dunbar, 1988) assume that the problem space for a task consists of two spaces: an hypothesis space and an experiment space. In hypothesis space, hypotheses about rules governing the task are generated, which can then be tested in experiment space. However, experiment space can be searched by applying the operators even without knowledge about the task. We predicted that people searching hypothesis space would learn more about the task. To test this claim, two experiments were performed in which subjects had to learn to control a system consisting of three input variables that had unknown links to three output variables. Subjects first explored the task, then they had to reach goal states for the output variables. In both experiments subjects were presented with an hypothesis about one of the links, which should foster search of hypothesis space. In Experiment 1, hypothesis instruction improved performance and we showed that it had a similar effect to a manipulation of goal specificity, suggesting that both factors improve learning by encouraging search in hypothesis space. In Experiment 2 subjects were given a correct hypothesis or an incorrect hypothesis. Both groups performed better than an appropriate control. Thus instructions that encourage hypothesis testing appear to improve learning in problem solving.

Introduction

Mayer's (1989) analysis of problem solving claimed that a problem solver applies representational processes to form a representation of a problem, then solution processes are applied to find the solution. However, Mayer notes that, as Duncker (1945) argued, representations may not be static and that the interaction of representational and solution processes may be the key to problem solving. But what processes form the most appropriate representations?

Dual Space Theory and Learning

One explanation for why some problem solvers learn more about a problem task than others is given in Klahr and Dunbar's (1988) theory of Scientific Discovery as Dual Search (SDDS). They propose that the problem space is separated into two spaces: an hypothesis space and an experiment space. Searching the hypothesis space requires formulating explicit hypotheses about the task, thus, the rules governing the task can be discovered. Searching the experiment space only requires applying the legal operators of the task to generate new problem states. According to SDDS theory a good representation of the task is gained if the problem solver searches both spaces interactively. Such problem solvers induce rules explicitly by searching the hypothesis space and they then test them through search of the experiment space. While search of experiment space is necessary for generating and testing hypotheses, a poor knowledge is gained if search of experiment space dominates. (However, such problem solvers may be the most successful if the rules are very hard to discover) The claim that problem solvers who formulate and test hypotheses have a better representation has found some support (Klahr & Dunbar, 1988; Recker, Govindaraj, & Vasandani, 1994). In particular, Klahr, Fay, and Dunbar (1993) found that subjects who generated hypotheses, even if incorrect, were more successful at solving a complex problem. However, these studies used a post-hoc classification of which problem space was searched. Therefore, to clearly show that using hypotheses improves learning, it is necessary to directly manipulate the likelihood of subjects generating and testing hypotheses.

Further support for the SDDS theory can be found in our own work (Vollmeyer, Burns, & Holyoak, in press). In these studies we used the theoretical framework of Simon and Lea (1974) to which Klahr and Dunbar (1988) also refer. Simon and Lea proposed that instance space (comparable with experiment space) is searched if problem solvers are focused on finding a solution for a specific goal, whereas rule space and instance space (rule space is comparable with hypothesis space) are searched if problem solvers are focused on learning the rules of the task. Therefore, we varied goal specificity by giving one group a non-specific goal, to learn as much as possible while exploring the problem task, then we tested their learning by giving them a goal state to reach; whereas another group explored the task with the same instruction, however in addition they were told at the start of the task the specific goal that they had to reach after the exploration phase. Consistent with the predictions, problem solvers with the non-specific goal had more knowledge about the rules governing the task and could apply the learned knowledge equally to two different goal states. Problem solvers with a specific goal learned less about the rules governing the task, but they could reach the specific goal they had been given at the start as well as the non-specific goal group. However, their performance declined when they were given a new goal to which such a solution path could not be readily transferred Rather than learn the rules, they may have learned a solution path while exploring the task. These results can be interpreted as supporting the claim that the non-specific goal group were more likely to search both spaces, whereas the specific goal group were more likely to only search the experiment space.

In the following two experiments we gave subjects an hypothesis about the structure of the task. With this manipulation we wanted to foster search of the hypothesis space. Under this condition they should gain more knowledge and consequently reach the goal state of the task more accurately. As the hypothesis also provided more information about the task to the subjects, in the second experiment we attempted to clearly establish that improved performance was due to search in hypothesis space and not just because more information was given.

Biology Lab: A Complex Problem Task

Vollmeyer, Burns, and Holyoak (in press) used a computer-driven problem task we called *biology lab* which was constructed using the shell DYNAMIS (Funke, 1991). This task was again used in the current experiments. In Vollmeyer et al. subjects had to control four output variables by varying four input variables, but in the first experiment we used a system with only three input and output variables. Subjects were presented with a cover story telling them that they were in a biology lab in which there were three species of sea animals in a tank (crabs, sea bass, lobster) and that their population could be manipulated by three factors (temperature, oxygen, current). The structure of the task (see Figure 1) was complex as one output (sea bass) was influenced by two inputs, and dynamic, as one output (lobster) had a decay (marked with a circle), resulting in the population decaying by 10 % each trial even if nothing was manipulated. As the decay was hard to understand we omitted this characteristic in the second experiment, which helps generalize our results to simpler systems.

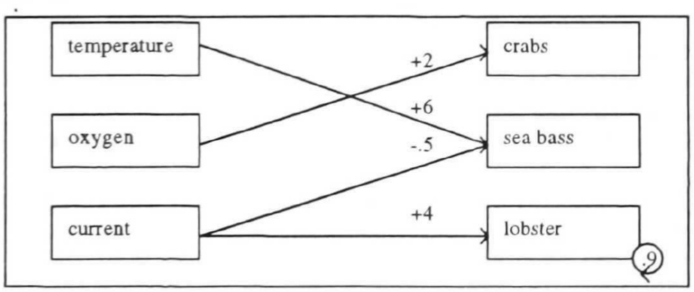

Figure 1: Biology lab system

To explore the task, subjects in Vollmeyer et al. (in press) were given a learning phase (three rounds of six trials on which numbers were entered for the inputs) and a solution round (six trials) at the end of which subjects had to reach a certain target amount for each output variable. Vollmeyer et al. showed that a good strategy for learning about the task was to vary only one input variable at a time. This strategy was given to all subjects in the current experiments in order to reduce their variance.

Experiment 1

In the first experiment we tested whether subjects given an hypothesis to test about a difficult relation in the problem task would learn more about the structure of the task and reach the goal state for the output variables more accurately than subjects given no hypothesis. We also manipulated goal specificity and predicted on the basis of Vollmeyer et al. (in press), that giving subjects a specific goal would produce similar effects to hypothesis instruction.

Method

Subjects. Sixty undergraduate students at the University of California, Los Angeles, participated for course credit.

Design. A 2x2 design was used with two levels of hypothesis-instruction (hypothesis-instructed vs. uninstructed) and goal specificity (specific vs. non-specific). Fifteen subjects were in each condition.

Procedure. The biology lab problem was presented with the underlying structure shown in Figure 1. Subjects had to learn about the problem in three rounds, each round with six trials and in the fourth round they were asked to reach a specific goal state (namely, 50 crabs, 900 lobsters, and 700 sea bass). Subjects in the specific goal condition were presented with these goal states right from the beginning, whereas the non-specific goal group saw these goal states at the beginning of the solution round for the first time.

Before starting, all subjects were instructed that the best strategy for exploring the task was to vary only one input variable at a time. In addition, the hypothesis-instructed group was told that a researcher believed that lobsters had a decay of 10% and that current had an influence in that each input to current is multiplied by four and then added to the lobster-population. Hypothesis-instructed subjects were told to test the hypothesis in order to determine if it was correct. Uninstructed groups received no hypothesis.

After each round of the learning phase (rounds 1-3) subjects completed a "structure diagram", which consisted of a diagram similar to the one in Figure 1, but with all links omitted. Subjects were instructed to draw links between variables that they believed affected each other, and to also assign directions (positive or negative) and weights indicating how strong they thought each influence was. After each input trial subjects had to predict the new values for each output variable that they thought would result from their inputs.

The entire experiment took an hour to complete.

Results

Dependent variables. Three dependent variables were analyzed which measured both knowledge and accuracy in reaching the goal states. (1) *Structure score*. The structure diagram was given after each of the three rounds of the learning phase. However, as the structure diagram after round 3 was most informative about subjects' knowledge at the end of the learning phase, only the structure score for this round is reported here. The knowledge indicated in this diagram was measured as the sum of the number of correct specifications of links, directions, and weights, adjusted with a correction for guessing (see Woodworth & Schlosberg, 1954, p. 700). (2) *Prediction error*. After each input trial during the learning phase subjects had to predict the population for each output variable. The absolute difference between the predicted number and the actual number for each of the three output variables was computed. As this measure produced a skewed distribution, the distribution was corrected by applying a logarithmic transformation. (3) *Solution error*. Solution error in reaching the goal state during round 4 was computed as the sum of the absolute differences between the new goal and the obtained number for each of the three output variables. Again, a logarithmic transformation had to be applied. Solution error was computed for each of the six trials that comprised round 4.

Preliminary analyses. The structure score and the sum of prediction errors (over three rounds) were measures of knowledge and should correlate, which was the case, $r = -.62$, $p < .001$. Having more knowledge should lead to lower solution errors. This was confirmed by the correlations for structure score and solution error, $r = -.48$, $p < .001$, and for prediction error and solution error, $r = .57$, $p < .001$.

Instructing subjects to test an hypothesis should help them gain more knowledge as measured by their structure diagram scores. In particular, if our manipulation was effective in getting subjects to test the given hypothesis then hypothesis-instructed subjects should be more likely to correctly report the links that were part of their hypothesis. We found this, as 18 out of 30 in the hypothesis-instructed groups correctly reported the decay factor for lobster compared to 4 out of 30 for uninstructed groups, $X^2(1) = 12.12$, $p < .001$. Hypothesis instruction also led more subjects to correctly specify the weight for the relation between current and lobster (also part of the hypothesis), 13 out of 30 compared to 1 out of 30 for uninstructed groups, $X^2(1) = 11.27$, $p < .001$.

Influence of hypothesis-instruction and goal specificity on learning. The hypothesis-instructed groups ($M = 1.55$) should have a higher structure score than the uninstructed groups ($M = 1.07$), which was the case, $F(1,56) = 4.58$, $p < .05$. Also, as predicted, the mean structure score for the non-specific goal groups are higher ($M = 1.56$) than that for the specific goal groups ($M = 1.06$), $F(1,56) = 5.13$, $p < .05$. There was no interaction between the factors $F < 1.0$.

Surprisingly, over all rounds of the learning phase there was no statistically significant effect of goal-specificity on prediction error, $F(1,56) = 1.69$, $p > .05$. However, there was an effect of hypothesis-instruction on prediction error, $F(1,56) = 8.92$, $p < .05$, and the interaction of output variable and hypothesis-instruction was significant, $F(2,112) = 9.31$, $p < .001$. Therefore we analyzed each output variable separately. For lobster, which was part of the hypothesis, a strong effect of hypothesis on predictions scores was found (see Table 1), $F(1,56) = 29.1$, $p < .05$. Although, for crabs and sea bass the difference was not significant, the means were in the expected direction (see Table 1).

For solution error, an effect of hypothesis-instruction was found, $F(1,56) = 8.20$, $p < .01$, but there was no interaction with output variable, $F < 1.0$. For our theory it is important to show that hypothesis-instruction helped performance on all output variables, not only the one the given hypothesis refers to, thus we analyzed each output variable further. As can be seen in Table 1, there was an effect on lobster, $F(1,56) = 14.35$, $p < .05$, and on crabs, $F(1,56) = 5.36$, $p < .05$. (Note that crabs were not referred to by the hypothesis) The hypothesis effect did not reach significance for sea bass, $F(1,56) = 2.36$, $p > .05$. As Vollmeyer et al. (in press) found there was no effect of goal specificity on solution error, $F < 1.0$.

Table 1. Means for hypothesis-instructed (H-I) vs. uninstructed (H-UI) subjects on prediction error and solution error, separated by output variable

output variables		prediction error	solution error
lobster	H-I	3.31	3.36
	H-UI	4.67	4.05
sea bass	H-I	3.25	3.94
	H-UI	3.61	4.49
crabs	H-I	1.55	1.83
	H-UI	1.88	2.86

Summary

Experiment 1 showed that having subjects test an hypothesis had an effect on learning. Hypothesis-instructed subjects learned more about the structure of the task and could predict the outcomes better than the uninstructed groups. All output variables were reached more accurately, not just the one referred to by the hypothesis, suggesting that having an hypothesis helps subjects learn about unrelated variables, perhaps through encouraging further hypothesis testing.

The results for goal-specificity replicated our previous experiment and showed goal specificity effects are generalizable to a different system. Subjects learned more about the structure if they had a non-specific goal, but they reached the goal states as well as the specific goal group, which already had experience in reaching the goal states. One surprising effect was that the goal groups did not differ on prediction error, which was another method for measuring knowledge. Perhaps this is because the specific goal groups were already focused on bringing about a specific state, which improved prediction, whereas the non-specific goal groups had more knowledge, but did not focus on reaching specific states. The lack of any interaction between goal specificity and hypothesis-instruction is interesting as it suggests that they may have their effects for similar reasons, that is, by encouraging search of hypothesis space.

Experiment 2

While Experiment 1 clearly showed that giving subjects an hypothesis improved both their knowledge and their performance, an alternative explanation is possible other than our claim that giving an hypothesis promotes search of hypothesis space. Because we gave subjects a correct hypothesis it is possible that they simply interpreted it as extra information and used it to help them control the system. Arguing against this possibility is the lack of an interaction on error-scores between output variable and hypothesis-instruction. However, if this alternative explanation is valid then giving subjects an hypothesis that is *incorrect* should eliminate the hypothesis effect. Klahr et al. (1993) found that subjects who generated incorrect hypotheses also performed better than those with no hypotheses, but they did not directly manipulate whether people generated hypotheses. Thus in Experiment 2 we tested whether giving subjects an incorrect hypothesis would help them to learn more about the biology lab task, as it would encourage search of hypothesis space. To do this we had three groups: *correct-hypothesis*, *incorrect-hypothesis*, and *link-only*. The correct-hypothesis group was instructed to test a correct hypothesis about a link and its weight. To reduce the usefulness of the information (but not the benefit of testing it) this hypothesis was about the simplest link, that between the input and output variable whose only link was to each other. Most subjects in previous experiments learned this particular link, so even if subjects assumed that this hypothesis was correct it would be of little use to them. The incorrect-hypothesis group was given a hypothesis about the same link, but they were told the wrong weight. The link-only group received the correct information that this same link existed, but no weight was suggested. Thus the link-only group had the same amount of correct information as the incorrect hypothesis group, but lacked the erroneous link information that made the incorrect-hypothesis group's hypothesis a complete one. We predicted that both the correct and incorrect hypothesis groups would perform better than the link-only group.

Other changes from Experiment 1 were that we changed the variable names and used a simpler biology lab system, as we dropped the decay link. Otherwise, the system was the same as that in Figure 1. These changes helped us generalize our results.

Method

Subjects. Two hundred and thirty-six students at the University of California, Los Angeles, participated.

Design. The experiment had three conditions, that is hypothesis-instruction was varied on three levels: correct-hypothesis; incorrect-hypothesis; and, link-only.

Procedure. The biology lab task was presented with the underlying structure shown in Figure 1, except that the decay link was omitted and the variable names were changed. The inputs, temperature, current and oxygen, became salt, carbon and lime, respectively. The outputs, crabs, sea bass and lobster, became oxygenation, chlorine concentration and temperature, respectively. As the task was easier subjects had only two rounds in the learning phase, during which they already knew the goal states for the learning round. From the beginning, all groups were given the goal state (namely, an oxygenation of 50, a chloride concentration of 700, and a temperature of 900). They had to reach this goal state in the third round. In the fourth round, the transfer round, a new goal state was given (namely, an oxygenation of 400, a chloride concentration of 700, and a temperature of 1000).

All subjects read the instructions explaining the task, and the same good strategy with which to explore the task as was given in Experiment 1. We had three levels of hypothesis-instruction: correct hypothesis, incorrect-hypothesis, and link-only information. In the correct-hypothesis group subjects were verbally and graphically presented with the hypothesis that lime could have an effect on oxygenation, that is, each input to carbon is multiplied by the weight 2.0 and then added to the oxygenation value. The incorrect-hypothesis group was told to test the same link, but the given weight (-5.0) was incorrect. Both hypothesis groups were instructed to test their hypothesis. The link-only group subjects were told that there could be a link between lime and oxygenation, but told no weight. In a previous experiment (Vollmeyer & Burns, under submission), we had found that giving links without direction or weights did not improve performance, thus this manipulation should be similar to giving subjects no hypothesis.

Because the system was simpler than that used in Experiment 1, only two rounds were given for the learning phase. After each round during this phase subjects completed the structure diagram, for which they were given detailed instructions for how to calculate weights. They again had to predict the outcomes for each output variable after each input trial. In order to encourage the hypothesis groups to test the hypothesis, these subjects were asked to indicate if they thought the hypothesis was correct by circling "Yes", "No", or "Don't know" after the end of each of the first two rounds. This question also measured if subjects given an hypothesis were able to determined it's validity.

The goal state for the solution round was presented to all subjects right from the beginning, however, they only had to reach that state in the third round. Therefore, subjects could decide, whether or not to focus on the specific goal. The results of Vollmeyer et al. (in press) suggested that giving subjects a specific goal from the start discourages them from testing hypotheses, thus giving a specific goal should decrease the a priori probability of subjects testing hypotheses. In round 4, subjects were presented with a new goal state which had to be reached. The entire experiment took an hour to complete.

Results

Dependent variables. The same three dependent variables as calculated in Experiment 1 were used, that is, structure score, prediction error, and solution error. Structure score and prediction error were measures of the knowledge subjects had of the rules governing the task. The solution error indicated whether subjects could apply their knowledge. As there was a transfer round in which a new goal state had to be obtained, a transfer error was calculated similar to the solution error. Transfer error measured how effectively the knowledge gained through trying to reach one goal state can be transferred to a new goal state.

Preliminary analyses. Again we checked whether our measures for learning were related, and whether our manipulation of hypothesis-instruction was effective. The sum of the prediction errors over three rounds and structure score on round 2 should be correlated as they both measure knowledge about the task, which was the case, $r = -.26$, $p < .001$, though this correlation was much lower than it was in Experiment 1. Having more knowledge should lead to lower solution and transfer errors, thus these measures should correlate, as we found; structure score and solution error: $r = -.61$, $p < .001$; prediction error and solution error: $r = .57$, $p < .001$; structure score and transfer error: $r = -.46$, $p < .001$; prediction error and transfer error: $r = .47$, $p < .001$. Transfer error and solution error were also correlated, $r = .81$, $p < .001$.

When asked if the given hypothesis was true, sixty-nine percent of responding subjects with the incorrect hypothesis believed it to be wrong, eight percent indicated it was correct, the rest were not sure. Eighty-one percent of responding subjects with the correct hypothesis believed it to be correct, twelve percent indicated incorrect, the rest were not sure. Thus most subjects appear to correctly test the hypothesis. As in Experiment 1 we analyzed subjects' success at finding the link that each group was given (lime to oxygenation), The difference between the three groups was significant, X^2 (2) = 6.38, $p < .05$. The correct hypothesis group indicated more often the correct weight (68 of 80) than the incorrect hypothesis group (58 of 78) and the link-only group (54 of 78).

Influence of hypothesis-instruction on learning. The correct- and incorrect-hypothesis groups should learn more about the structure of the task. However, they do not differ on the structure score, $F < 1.0$. Together with the low correlation between prediction error and structure scores, this suggest that structure score may not be a good measure of knowledge for a system as simple as this.

As predicted though, the prediction error over the three rounds showed an effect of hypothesis condition, $F(2,233) = 5.36$, $p < .01$ (see Table 2). Link-only prediction errors are higher than either those of the incorrect hypothesis, $F(1,154) = 7.14$, $p < .01$, or correct hypothesis groups, $F(1,156) = 8.21$, $p <. 01$. Thus it appears that it is more important that a hypothesis be given, rather than whether the hypothesis is correct.

If subjects perform better just because of the amount of correct information they are given then the predictions for the output variable oxygenation should have been best for the correct hypothesis group, while there should have been little effect on other variables. According to our theory, all output variables should be better predicted, if a hypothesis was given, no matter whether the hypothesis is correct or incorrect. A significant interaction between hypothesis-instruction and output variable allowed us to analyze each single output variable, $F(4,466) = 4.73$, $p < .001$. Table 3 shows the means for the three experimental groups. For the crucial comparison, that is link-only vs. incorrect hypothesis, we found significant differences for the prediction error of chloride concentration, $F(1,154) = 7.96$, $p < .01$, and for the prediction error of temperature, $F(1,154) = 9.30$, $p < .01$, however not for the prediction error of oxygenation, $F(1,154) = 2.53$, $p > .05$, the output variable for which the hypothesis was given. All differences between correct and incorrect hypothesis groups were not significant.

Table 2. Means of hypothesis-instruction on the dependent variables

	prediction error	solution error	transfer error
link-only	2.82	2.61	2.61
incorrect hypothesis	2.18	2.23	2.19
correct hypothesis	2.11	1.87	1.83

Table 3. Means of hypothesis instruction on the prediction error of the output variables

	chloride concentration	temperature	oxygenation
link-only	3.51	2.90	2.06
incorrect hypothesis	2.83	1.99	1.73
correct hypothesis	2.82	2.02	1.49

We analyzed whether hypothesis-instruction had an influence on solution and transfer error. Across both errors there was an effect of hypothesis-instruction, $F(2,233) = 3.97$, $p < .05$. As can be seen in Table 2, there was a difference between the link-only group and the correct hypothesis group. However, the expected differences between the link-only group and the incorrect hypothesis group are not significant (solution error: $F(1,154) = 1.86$, $p > .05$; transfer error: $F(1,154) = 1.93$, $p > .05$). As in Experiment 1 there is no significant interaction between errors and output variables, $F(2,466) = .49$, which is important for our point that giving an hypothesis does not only assist performance on the output variable on which the information was given.

Summary

Experiment 2 showed that even hypothesis-instruction with an incorrect hypothesis can improve performance, even when compared to a group given the same valid information but without as extensive an hypothesis to test. This is consistent with the claim that hypothesis-instruction not only provides information, but also leads to another type of processing, that is search in hypothesis space through the generation of hypotheses.

Discussion

Our aim was to find empirical evidence addressing why some people form a good representation during learning of a problem task, while other people have difficulties in finding a solution. Our theoretical explanation was based on dual space models, such as SDDS, that assumes that searching the hypothesis space by generating hypotheses helps learning. Goal specificity (Experiment 1) as well as hypothesis-instruction (both experiments) seem to be factors that have an influence on the choice of how to represent the task.

The results of the two experiments are not always statistically significant on all of the dependent variables, but the pattern on these measures is always as expected. Even if we changed the task from a dynamic (Experiment 1) to a simpler task (Experiment 2), or gave an hypothesis about a simple or complex link, the influence of hypothesis-instruction was consistent. However, changing the task to a simpler system had consequences on the performance. One consequence was that subjects on average learned more about the structure of the simpler task (M = 2.13) than about a dynamic task (M = 1.32). As most of the people seem to learn the simple task, the structure score does not differentiate anymore. This explains why hypothesis-instruction had an effect on structure score in the first, but not in the second experiment. Predicting the outcome of a manipulation of the input variables demonstrated clear effects of hypothesis-instruction in both experiments. With a better representation of the task subjects generating hypotheses are able to reach given goal states more accurately. Therefore, we regard the results as encouraging evidence that problem-solving can be most effective when the problem space is represented as a dual space. Such a representation appears to be encouraged if subjects are given an hypothesis, even an incorrect one.

Other recent studies can be also be interpreted as evidence that search in hypothesis space improves performance. Chi, de Leeuw, Chui, and LaVancher (1994) found that instructing subjects to generate explanations of a text while they read it improved learning of its content, despite a fourth of these self-explanation being incorrect. Self-explanations may be like hypotheses and assist search of hypothesis space.

Klahr (1994) argues that cognitive psychology and machine learning approaches to scientific discovery have converged towards dual-space theories. Machine learning has generally taken a highly data driven approach, which is more akin to search of experiment space. But this space needs to be limited in order to make such search tractable, and domain knowledge provided by testing hypothesis may provide these constraints. Case-based reasoning would seem to be an extreme form of search of the experiment space. But if cases constitute experiment space (past experiments), domain knowledge provides the hypothesis space. That case-based reasoning needs to consider more than just cases has been argued by Kolodner (1994).

Our results demonstrate the importance of finding general rules for a problem instead of simply using a already learned solution path, thus they support a dual-space search approach to reasoning

Acknowledgements

This research was supported by DFG Grant Vo 514/1 to Regina Vollmeyer and NSF Grant SBR-9310614.

References

Chi, M. T. H., de Leeuw, N., Chiu, M-H, & LaVancher, C. (1994). Eliciting self-explanation improves understanding. *Cognitive Science, 18,* 439-477.

Duncker, K. (1945). On problem-solving. *Psychological Monographs, 58,* 1-113.

Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 185-222). Hillsdale, NJ: Erlbaum.

Klahr, D. (1994). Children, adults, and machines as discovery systems, *Machine Learning, 14,* 313-320.

Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12,* 1-55.

Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology, 25,* 111-146.

Kolodner, J. L. (1994). Understanding creativity: A case-based approach. In S. Wess, Althoff, K-D, & Richter, M. M. (Eds.), *Topics in case-based reasoning* (pp 3-20). Berlin: Springer-Velag.

Mayer, R. E. (1989). Human nonadversary problem solving. In K. J. Gilhooly (Ed.), Human and machine problem solving (pp. 39-56). New York: Plenum Press.

Recker, M. M., Govindaraj, T. & Vasandani, V. (1994). Troubleshooting in a complex, dynamical domain. In A. Ram & K. Eiselt (Eds.) *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 739-744). Hillsdale, NJ: Erlbaum.

Simon, H. A. & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. W. Gregg (Ed.), *Knowledge and cognition* (pp. 105-127). Hillsdale, NJ: Erlbaum.

Vollmeyer, R. & Burns, B. (under submission). Hypothesentesten beim Erlernen eines Problems.[Hypothesis testing in learning a problem task]

Vollmeyer, R., Burns, B. D. & Holyoak, K. J. (in press). The impact of goal specificity and systematicity of strategies on the acquisition of problem solving. *Cognitive Science*.

Woodworth, R. S. & Schlosberg, H. (1954). *Experimental psychology*. New York: Holt, Rinehart, and Winston.